AMEA Nəsimi adına Dilçilik İnstitutu,
Dilçilik araşdırmaları, 2026, №1 *ISSN: 2664-5432*
Səhifə: 178-183
https://doi.org/10.59849/2664-5432.2026.1.178

Gülnar SAYILOVA*

AZƏRBAYCAN DİLİNDƏ NİTQ MATERİALLARININ ARXİVLƏŞDİRİLMƏSİ PROSESİ: ELMİ PRİNSİPLƏR VƏ MÜASİR YANAŞMALAR

XÜLASƏ

Məqalədə nitq materiallarının elmi metodlarla toplanılması, rəqəmsallaşdırılması və uzunmüddətli qorunması prossesi əvvəldən axıradək izahlı formada öz əksini tapmışdır. Nitq materiallarının məqsədyönlü şəkildə toplanması və sistemli formada işlənməsi dil tədqiqatlarının etibarlılığını təmin edən başlıca şərtlərdəndir. Burada şifahi nitq nümunələrinin audioyazı və transkripsiya vasitəsilə işlənib hazırlanmasının prinsipləri, eləcə də fonetik və leksik tədqiqatlar üçün onun yaratdığı imkanlar diqqətə çatdırılır. Eyni zamanda, toplanmış nitq materiallarının sistemləşdirilməsi zamanı mövzu, areal, danışan şəxsin sosial xüsusiyyətləri və nitqin funksional tipi kimi meyarların nəzərə alınmasının vacibliyi vurğulanır. Bu meyarlar əsasında işlənmiş nitq materiallarının regional xüsusiyyətlərin müqayisəli təhlili və müasir dil proseslərinin izlənilməsi üçün zəngin elmi baza yaratdığı göstərilir.

Tədqiqatda **ELAN** proqramının bu prosesdə mühüm rolu və nitq materiallarının saxlanılmasında əsas vasitələrdən biri olduğu göstərilmişdir. Nəticə etibarilə, nitq materiallarının arxivləşdirilməsi Azərbaycan ədəbi dili, Azərbaycan dilinin dialektləri, eləcə də Azərbaycanda yaşayan müxtəlif xalqların nitq materialının qorunması, gələcək nəsillərə çatdırılması və inkişafı üçün mühüm istiqamət kimi dəyərləndirilir.

Açar sözlər: *Nitq materiallarının arxivləşdirilməsi, rəqəmsallaşdırma, ELAN proqramı, dilin qorunması.*

Azərbaycan dilində nitq materiallarının arxivləşdirilməsi müasir dilçilik, fonetika, dialektologiya, nitq texnologiyaları və mədəni irsin qorunması üçün böyük əhəmiyyət daşıyan prosesdir. Nitq materiallarının arxivləşdirilməsinin əsas məqsədi dillərin və şivələrin itmə riskini azaltmaq, bu materialların uzunmüddətli qorunması üçün şərait yaratmaqdır.

Müasir dövrdə informasiya texnologiyalarının sürətli inkişafı dilçilik sahəsində də yeni metod və vasitələrin yaranmasına səbəb olmuşdur. Xüsusilə nitqin, danışıq nümunələrinin və audiovizual materialların tədqiqi zamanı xüsusi proqram təminatlarından geniş istifadə edilir.

* AMEA Nəsimi adına Dilçilik İnstitutu, Eksperimental Fonetik Araşdırmalar Laboratoriyası. Email: sgulnar0404@gmail.com

Bu vasitələrdən biri də **ELAN** proqramıdır. Bu tətbiq dilçilik, etnoqrafiya, folklorşünaslıq, antropologiya və kommunikasiya tədqiqatlarında audiovizual materialların annotasiyası və təhlili üçün istifadə olunan peşəkar proqram təminatıdır. Proqram ilk olaraq Niderlandın Maks Plank Psixolinqvistika İnstitutu tərəfindən hazırlanmışdır və hazırda dünya miqyasında elmi araşdırmalarda geniş tətbiq olunur.

ELAN proqramı xüsusilə şifahi nitqin, dialektlərin, jest və mimikaların, həmçinin multimodal ünsiyyət vasitələrinin öyrənilməsi üçün əlverişli imkanlar yaradır. Bunun əsas üstünlüyü audio və video materialların zaman oxu üzrə dəqiq şəkildə işarələnməsi və müxtəlif qatlarda annotasiya edilməsidir. Bu xüsusiyyət tədqiqatçıya nitq materialını sistemli şəkildə təhlil etməyə imkan verir. [Austin, P.,2017,s.29]

Proqramın mühüm cəhətlərindən biri də məlumatların ixrac və inteqrasiya imkanlarıdır. ELAN proqramında yaradılmış annotasiyalar XML və digər formatlarda saxlanıla bilir. Bu isə məlumatların başqa proqram təminatlarına ötürülməsini və müxtəlif elmi platformalarda istifadəsini mümkün edir. Nəticədə tədqiqat materiallarının arxivləşdirilməsi və gələcək nəsillər üçün qorunması təmin olunur.

ELAN proqramı əsasında Nitq materialının arxivləşdirilməsi zamanı arxitekturanın formalaşması vacib proseslərdəndir. Bu zaman ilkin mərhələdə nitq nümunələrinin sistemli və ardıcıl şəkildə aşağıdakı mənbələr əsasında toplanması nəzərdə tutulur:

1.Dialekt və şivə materialları;
2. Mədəni irs nümunələri;
3.Akademik nitq nümunələri;
4.Müasir mənbələr.

Nitq materiallarının arxivləşdirilməsi zamanı diqqətəlayiq çəkiyə malik material kimi dialekt və şivə materialları çıxış edir. Bu tipli nümunələrin toplanması zamanı, əsasən bir neçə xüsusiyyət nəzərə alınmalıdır ki, bunlardan daha çox əhəmiyyət kəsb edəni regional fonetik xüsusiyyətlər və leksik fərqlərdir. Bununla yanaşı, qeyd olunan məqamda həm də dilin tarixi qatlarını əks etdirən məlumatlar diqqət çəkir.

Arxivləşdirmə prossesində dialekt və şivələrdən sonra ikinci əsas mənbə kimi mədəni irs nümunələrini göstərmək olar. Mədəni irs nümunələri dedikdə, daha çox xalq nağılları, aşıq yaradıcılığı, folklor mətnləri nəzərdə tutulur. Eyni zamanda adət-ənənələrə dair məlumatlar da mədəni irs nümunələrində yer alan əsas materiallardandır.

Akademik nitq materiallarının toplanmasında məqsəd dilin ədəbi normalarını müəyyənləşdirmək və dilçilik tədqiqatları üçün əsaslı şəkildə baza yaratmaqdır. Belə nümunələrə alim və professorların müsahibələri, elmi müzakirələr və tarixi yaddaş üçün mühün sayılan şəxslərin səsli nitq nümunələri daxildir.

Nitq materiallarının toplanmasında sonuncu vasitə kimi müasir mənbələri, yəni radio və televiziya çıxışlarını, rəqəmsal platformalardakı nitq nümunələrini və

məktəblərdə, universitetlərdə aparılan müsahibələri göstərmək olar. Belə mənbələr nitq arxivləşdirilməsi zamanı gələcəkdə aparılacaq tədqiqatlar üçün etibarlı və geniş verilənlər bazası rolunu oynayır [Baldwin, 2017: s.27]

Göstərilən mənbələr əsasında toplanan materiallar müasir texnologiyalar vasitəsilə rəqəmsal formaya keçirilir. İlk öncə səsin rəqəmsal formaya keçirilməsi prossesi başlayır və bu prosses aşağıdakı kimidir:

1.Səsin rəqəmsal formaya keçirilməsi:

- WAV, FLAC və ya MP3 formatlarında saxlanılma;
- 44.1kHz və ya 48 kHz səs tezliyi ilə yazı;

2.Səsin təmizlənməsi:

- Şəkilləndirmə;
- Tezliyin balanslaşdırılması;
- Qeyri-arzuolunan səslərin çıxarılması (küydən təmizlənmə);

3.Transkripsiya:

- Fonetik transkripsiya (Azərbaycan fonetik sistemi ilə);
- Orfoqrafik transkripsiya (Ədəbi dil normalarında);

4.Meta-məlumatların hazırlanması:

- Danışanın yaşı, cinsi, təhsili;
- Dialekt arealı;
- Yazıya alınma tarixi;
- Nitq mövzusu;

Yuxarıda verilən ardıcıllıqdan sonra isə arxivləşdirilmə mərhələsi başlayır. Bu prosses aşağıda göstərilən formada və bir neçə bəndlə özünü göstərir.

1.Rəqəmsal arxivlərin yaradılması:

- Bulud sistemləri;
- Server əsaslı məlumat bazaları;
- Rəqəmsal kitabxanalar;

2.Kataloqlaşdırma:

- Materiallar dil, bölgə, mətnin tipi, janr və digər göstəricilərə görə qruplaşdırılır.

3.Ehtiyat nüsxələrin saxlanılması:

- Hər bir material ən az 2-3 fərqli diskdə və serverdə saxlanılmalıdır.

Hazırlanan nitq materialları, **ELAN** proqramında sistemləşdirilir və seçilmiş əməliyyat həyata keçirilir.[https://commonvoice.mozilla.org/ru]

ELAN proqramı müasir nitq texnologiyaları və dil korpuslarının emalı sahəsində geniş istifadə olunan annotasiya sistemlərindən biridir. Bu proqram nitq materiallarının sistemləşdirilməsi, emalı və müxtəlif səviyyələrdə təhlil olunması üçün funksional imkanlar təqdim edir. Proqram xüsusilə audiovizual materialların annotasiyası zamanı istifadəçilərə çevik və çoxşaxəli idarəetmə mexanizmi təqdim etməsi ilə seçilir. Proqramın əsas məqsədi nitq məlumatlarının strukturlaşdırılmış şəkildə saxlanılması və tədqiqat prosesində daha səmərəli istifadəsinin təmin

edilməsidir. İstifadəçi annotasiyaları bir neçə görünüş formasında izlənilə bilir və onlar media qrafiki ilə sinxronlaşdırılır. Bu xüsusiyyət audiovizual materialların daha dəqiq təhlilinə şərait yaradır və nitq vahidlərinin zaman intervalı üzrə düzgün müəyyənləşdirilməsinə imkan verir. Sinxronlaşdırma prosesi tədqiqatçıya audio və video materiallarla paralel işləmək imkanı yaradaraq məlumatların interpretasiyasını asanlaşdırır.

Proqramın digər üstün cəhətlərindən biri çoxsəviyyəli və təksəviyyəli iyerarxik strukturların yaradılmasını dəstəkləməsidir. Bu imkan müxtəlif annotasiya qatlarının formalaşdırılmasına şərait yaradır və nitq materiallarının fonetik, morfoloji, sintaktik və semantik baxımdan ayrıca təhlil edilməsinə kömək edir. İyerarxik quruluş tədqiqat prosesində məlumatların sistemli şəkildə təşkilinə və əlaqəli elementlərin daha rahat müəyyən edilməsinə xidmət edir.

ELAN proqramı nəzarət olunan lüğətlərin istifadəsini dəstəkləyərək terminoloji vahidlərin standartlaşdırılmasına şərait yaradır. Bu xüsusiyyət elmi tədqiqatlarda vahid terminologiyanın qorunması baxımından da mühüm əhəmiyyət daşıyır. Eyni zamanda, proqram bir annotasiya sənədinə bir neçə video faylın əlavə edilməsinə imkan verir ki, bu da paralel audiovizual materialların müqayisəli təhlilini asanlaşdırır.

Media dəstəyi baxımından proqram müxtəlif platformalarla uyğun şəkildə işləyir. Bu Windows Media Player, QuickTime və VLC kimi media çərçivələri üzərində qurularaq audio və video faylların effektiv idarə olunmasını təmin edir. Audio və video formatlarının dəstəklənməsi əməliyyat sistemindən asılı olsa da, proqram yüksək performanslı media oxunmasına nail olur. Bu isə böyük həcmli audiovizual materiallarla işləyən tədqiqatçılar üçün mühüm texniki üstünlük hesab edilir. [Hiippala,2021: s.657].

Nəticə etibarilə, ELAN proqramı nitq materiallarının annotasiyası və təhlili sahəsində funksional imkanları ilə seçilən müasir proqram təminatıdır. Onun təqdim etdiyi çoxsəviyyəli struktur, media sinxronizasiyası və geniş texniki dəstək dilçilik, kommunikasiya və multimodal tədqiqatların daha səmərəli aparılmasına mühüm töhfə verir.

ELAN proqramı vasitəsilə aparılan multimodal annotasiya prosesi nitq materiallarının müxtəlif qatlar üzrə sistemli şəkildə (video, səs və mətn qatları üzrə) transkripsiya olunmasını əyani şəkildə nümayiş etdirir. Proqramın təqdim etdiyi səs dalğası qrafikləri (PRAAT) nitqin intensivlik, pauza və ritm xüsusiyyətlərini müəyyən etməyə eyni zamanda səs,mətn və zaman göstəricilərinin paralel şəkildə işlənməsinə imkan verir [Xu, J. & Li,T.2022; Lin, Y.2020]

ELAN mühitində aparılan bu cür annotasiya prosesi nitq vahidlərinin kontekst daxilində sistemli şəkildə təsnifinə və elmi təhlilinə şərait yaradır. Belə texniki vasitələr dialektoloji, fonetik və eksperimental dilçilik araşdırmalarında etibarlı empirik materialın formalaşdırılmasında mühüm rol oynayır. Annotasiya olunmuş nitq materiallarının sistemli şəkildə saxlanılması və strukturlaşdırılması tədqiqatçılar üçün materialların daha operativ təhlilinə imkan yaradır.

Eyni zamanda rəqəmsal nitq bazalarının yaradılması müasir dil texnologiyalarının inkişafına, xüsusilə avtomatik nitq tanınmasına, dilin sağlam şəkildə qorunması və gələcək nəsillərə çatdırılması üçün geniş imkanlara yol açır. Nitq vahidlərinin zaman oxu üzrə seqmentləşdirilməsi fonetik dəyişmələrin və artikulyasiya xüsusiyyətlərinin daha aydın müşahidə olunmasına kömək edir. Sistemli şəkildə qurulan nitq arxivləri yalnız dilçilik tədqiqatlarında deyil, həm də süni intellekt, təhsil və mədəni irsin öyrənilməsi sahələrinə mühüm töhfə verir [Snaith,2021; Tian,2022; Wang, 2022]. **ELAN** proqramı vasitəsilə hazırlanan bu materiallar gələcəkdə Azərbaycan dilinin milli korpusuna əlavə edilə və müxtəlif layihələrdə istifadə oluna bilər.

ƏDƏBİYYAT

1. Austin Peter K. 2017. Language documentation and legacy text materials. Asian and African Languages and Linguistics 11, 23-44.
2. Baldwın, Daryl. Language Documentation & Legacy Text Materials. Asian and African Languages and Linguistics. n. 11. p. 23-44. 2017.
3. Common Voice Project by Mozill. https://commonvoice.mozilla.org/ru.
4. Hiippala, T. AI2D-RST: a multimodal corpus of 1000 primary school science diagrams. Lang. Resour. Eval.55 (3), 661–688 (2021).
5. Xu, J. & Li, T. Application of multimodal NLP instruction combined with speech recognition in oral english practice. Mobile Inform. Syst. 1 (2022) 2262696. (2022).
6. Lin, Y. A unified framework for multilingual speech recognition in air traffic control systems. IEEE Trans. Neural Netw. Learn. Syst.32 (8), 3608–3620 (2020).
7. Snaith, M. A multimodal corpus of simulated consultations between a patient and multiple healthcare professionals. Lang. Resour. Eval.55 (4), 1077–1092 (2021).
8. Tian, M. Construction of computer English corpus assisted by internet of things information perception and interaction technology. Comput. Intell. Neurosci. 1 (2022) 6803802. (2022).
9. Wang, M., Chen, J., Zhang, X. L. & Rahardja, S. End-to-end multi-modal speech recognition on an air and bone conducted speech corpus. IEEE/ACM Trans. Audio Speech Lang. Process.31, 513–524 (2022).

PROCESS OF ARCHIVING SPEECH MATERIALS IN THE AZERBAIJAN LANGUAGE: SCIENTIFIC PRINCIPLES AND MODERN APPROACHES

SUMMARY

The article describes the process of collecting, digitizing, archiving and long-term preservation of speech materials using scientific methods from beginning to end in an explanatory form. It is shown that the announcement program plays an important role in the archiving process and is one of the main tools in archiving speech materials. As a result, archiving speech materials is considered an important direction for the scientific preservation and development of the Azerbaijani language. It has been shown that the

announcement program plays an important role in this process and is one of the main tools for preserving speech materials. Consequently, archiving speech materials is considered an important direction for the preservation, transmission and development of the Azerbaijani language to future generations.

It has been shown that the announcement program plays an important role in this process and is one of the main tools for preserving speech materials. Consequently, archiving speech materials is considered an important direction for the preservation, transmission and development of the Azerbaijani language to future generations.

Keywords: *Archiving of speech materials, digitization, ELAN program, language preservation.*

ПРОЦЕСС АРХИВИРОВАНИЯ РЕЧЕВЫХ МАТЕРИАЛОВ НА АЗЕРБАЙДЖАНСКОМ ЯЗЫКЕ

РЕЗЮМЕ

В статье в пояснительной форме описывается процесс сбора, оцифровки и долговременного сохранения речевых материалов с использованием научных методов от начала до конца. Целенаправленный сбор и систематическая обработка речевых материалов являются одними из главных условий обеспечения достоверности лингвистических исследований. Обращается внимание на возможности, создаваемые подготовкой образцов устной речи для научного распространения посредством аудиозаписи и транскрипции для фонетических и лексических исследований. В статье подчеркивается важность учета таких критериев, как тематика, регион, социальные характеристики говорящего и функциональный тип речи, при систематизации собранных речевых материалов. Показано, что речевые материалы, обработанные на основе этих критериев, создают богатую научную базу для сравнительного анализа региональных особенностей и мониторинга современных языковых процессов.

Доказано, что программа объявлений играет важную роль в этом процессе и является одним из основных инструментов сохранения речевого материала. Следовательно, архивирование речевого материала рассматривается как важное направление для сохранения, передачи и развития азербайджанского языка будущим поколениям.

Ключевые слова: *Архивирование речевых материалов, оцифровка, программа ELAN, сохранение языка.*